



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/672,248	09/29/2003	Monika Henzinger	0026-0043	8702
44989 7590 01/09/2008 HARRITY SNYDER, LLP 11350 Random Hills Road SUITE 600 FAIRFAX, VA 22030			EXAMINER RUTLEDGE, AMELIA L	
			ART UNIT 2176	PAPER NUMBER
			MAIL DATE 01/09/2008	DELIVERY MODE PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.



UNITED STATES PATENT AND TRADEMARK OFFICE

Commissioner for Patents
United States Patent and Trademark Office
P.O. Box 1450
Alexandria, VA 22313-1450
www.uspto.gov

MAILED

JAN 09 2008

Technology Center 2100

**BEFORE THE BOARD OF PATENT APPEALS
AND INTERFERENCES**

Application Number: 10/672,248
Filing Date: September 29, 2003
Appellant(s): HENZINGER ET AL.

Viktor Simkovic
For Appellant

EXAMINER'S ANSWER

This is in response to the appeal brief filed October 26, 2007 appealing from the Office action mailed January 18, 2007.

(1) Real Party in Interest

A statement identifying by name the real party in interest is contained in the brief.

(2) Related Appeals and Interferences

The examiner is not aware of any related appeals, interferences, or judicial proceedings which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of Claims

The statement of the status of claims contained in the brief is correct.

(4) Status of Amendments After Final

The appellant's statement of the status of amendments after final rejection contained in the brief is correct.

(5) Summary of Claimed Subject Matter

The summary of claimed subject matter contained in the brief is correct.

(6) Grounds of Rejection to be Reviewed on Appeal

The appellant's statement of the grounds of rejection to be reviewed on appeal is correct.

(7) Claims Appendix

The copy of the appealed claims contained in the Appendix to the brief is correct.

(8) Evidence Relied Upon

Galai et al., PCT Application filed August 2002, International Publication Number WO 03/017023 A2, published February 2003.

6,665,658

DACOSTA et al.

12-2003

(9) Grounds of Rejection

The following ground(s) of rejection are applicable to the appealed claims:

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1-8, 10, 12-21, 23-26, and 28-30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai et al. (hereinafter "Galai"), PCT Application filed August 2002, International Publication Number WO 03/017023 A2, published February 2003, in view of DaCosta et al. (hereinafter "DaCosta"), U.S. Patent No. 6,665,658, issued December 2003.

Independent claim 1 cites: *A method for crawling documents comprising:*
receiving a uniform resource locator (URL);
receiving at least two different copies of a document associated with the URL;
and determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold.

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the Web page includes one or more links with the complete URL, as for a session ID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical.

Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21).

Galai teaches comparing a portion of the URLs that change between the two copies of the document and determining similarity based on a predetermined value of the portion of the URLs that change (p. 27, l. 6- p. 28, l. 21; especially p. 28, l. 11-21), since Galai teaches automatically determining the redundant parameter of the URL comparisons of divisible subunits of the URL and then using that parameter, or URL portion, as a basis of comparison to other URLs. Galai teaches determining a similarity level which is the likelihood of two web pages to have the same content, and if this value exceeds a certain threshold, then the URL portion, i.e., the subunit of the URL, is determined to be redundant (p. 27, l. 6-22; Fig. 5). Galai teaches using the threshold to determine similarity between URLs for purposes of crawling and indexing a web page for a search engine (p. 26, l. 19-p. 27, l. 5), and it would have been obvious to one of ordinary skill in the art at the time of the invention that the same threshold could be used to determine the difference between URLs for web pages, since the difference threshold would have been the inverse or opposite of the similarity threshold number.

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a

software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

Regarding dependent claim 2, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

Regarding dependent claim 3, Galai teaches a method of normalizing the URL in order to index substantially similar web pages only once (p. 20, l. 10-23); i.e., *comparison of the clean, or normalized URL to a set of clean URLs that represent previously crawled URLs*; since Galai teaches a process for detecting redundant parameters in URLs with the same structure, executed once per URL structure, and

then applied and executed for application to each URL with the same structure (p. 21, l.21-p. 22, l. 5).

Regarding dependent claims 4-6, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20). Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are determined automatically by comparing the URLs for redundancy and normalizing them.

Regarding dependent claim 7, Galai teaches a process for detecting redundant parameters in URLs with the same structure, executed once per URL structure, and then applied and executed for application to each URL with the same structure (p. 21, l.21-p. 22, l. 5), compare to *receiving the URL as a URL from a previously crawled web document*.

Regarding dependent claim 8, Galai teaches crawling the URL when the URL is determined to not already have been crawled (p. 24, l. 9-15).

Independent claim 10 cites: *A method for identifying web sites that use session identifiers comprising: downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; comparing the extracted URLs of the two different copies of the document; and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.*

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches that the web page is retrieved again using the reduced URL, i.e., downloading at least two different copies of at least one page from a web site. Galai teaches the comparison of URLs within the documents where the Web page includes one or more links with the complete URL, as for a session ID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21) when the comparison indicates that at least a portion of the URLs change between the two different copies.

Galai teaches comparing a portion of the URLs that change between the two copies of the document and determining similarity based on a predetermined value of the portion of the URLs that change (p. 27, l. 6- p. 28, l. 21; especially p. 28, l. 11-21), since Galai teaches automatically determining the redundant parameter by URL comparison and then using that parameter as a basis of comparison to other URLs.

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL, Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches

Art Unit: 2176

a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

Regarding dependent claim 12, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

Regarding dependent claim 13, claim 13 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

Regarding dependent claim 14, Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are generated automatically by the method of comparing the URLs for redundancy and normalizing them.

Independent claim 15 cites: *A device comprising: a spider component configured to crawl web documents associated with at least one web site; and a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of uniform resource locators (URLS) that change between different copies of at least one web document downloaded from the web site.*

Galai teaches a method of indexing dynamic web pages for a search engine. The search engine consists of a spider and repository (p. 4, l. 3-19). Galai teaches a method for normalizing the URL of a document to index substantially similar Web pages only once (p. 20, l. 10-20). Galai teaches comparing a Web page with a second retrieved web page with reduced parameters, i.e., any divisible subunit of the URL (p. 20, l. 10-20). Galai teaches the comparison of URLs within the document where the Web page includes one or more links with the complete URL, as for a session ID (p. 20, l. 21-p. 21, l. 9), resulting in two web pages which are similar in content but not identical. Galai teaches detecting the change between the two different copies of the document (p. 21, l. 1-8; p. 27, l. 14-p. 28, l. 21).

While Galai teaches a comparison of URLs for redundant parameters, which would include session identifiers since session identifiers are parameters within a URL,

Galai does not explicitly teach that the URLs are compared for the specific purpose of determining whether the web site uses session identifiers. However, DaCosta teaches a method for crawling documents in a dynamic website, with a database for storing and identifying session identifiers URLs, and an application program for controlling a software agent (Col. 4, l. 41-Col. 5, l. 23). DaCosta teaches the analysis of URLs and headers containing cookie data to determine if a web site uses session IDs (Col. 6, l. 21-40). It was notoriously well known in the art at the time of the invention that session data for a web site and/or document could be contained in either the URL string, or in a cookie.

Both DaCosta and Galai are directed toward methods for crawling web documents and tracking state and session information for web documents. It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the method of indexing web pages for a search engine by removing redundant pages by comparing URL parameters taught by Galai, with the means of identifying session identifiers by comparing URL data and cookie data taught by DaCosta, so that Galai would have the benefit of identifying session information for a web site whether the session information were contained in the URL string or in the cookie, in order to remove redundant pages from both configurations (URL string or cookie) of dynamic web sites.

Regarding dependent claim 16-17, Galai teaches a spider to download content from a network and a component of the autonomous software search program to extract URLs from the downloaded content (p. 26, l. 19-p. 27, l. 5) compare to *fetch component*

configured to download content from a network; and a content manager configured to extract URLs from the downloaded content.

Regarding dependent claim 18, claim 18 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

Regarding dependent claim 19, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20).

Regarding dependent claim 20, Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are generated automatically by comparing the URLs for redundancy and normalizing them.

Regarding independent claim 21, claim 21 is directed toward the device used for implementing the method as claimed in claim 10, and is rejected along the same rationale.

Regarding dependent claim 23, claim 23 reflects substantially similar subject matter as claimed in dependent claim 12, and is rejected along the same rationale.

Regarding dependent claim 24, claim 24 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

Regarding dependent claim 25, claim 25 reflects substantially similar subject matter as claimed in dependent claim 14, and is rejected along the same rationale.

Regarding independent claim 26, claim 26 is directed toward the computer-readable medium containing programming instruction for executing the method as claimed in claim 10, and is rejected along the same rationale.

Regarding dependent claim 28, claim 28 reflects substantially similar subject matter as claimed in dependent claim 12, and is rejected along the same rationale.

Regarding dependent claim 29, claim 29 reflects substantially similar subject matter as claimed in dependent claim 2, and is rejected along the same rationale.

Regarding dependent claim 30, claim 30 reflects substantially similar subject matter as claimed in dependent claim 14, and is rejected along the same rationale.

(10) Response to Argument

Beginning on page 5 of the Appeal Brief (hereinafter "the Brief"), Appellant argues the following issues which are addressed below.

"A. The Rejection under 35 U.S.C. § 103(a) Based on Galai and DaCosta Should be Reversed." (pages 5-19 of the Brief).

"1. Claims 1-3 and 5-8." (pages 6-12 of the Brief).

The examiner maintains that the combination of Galai and DaCosta discloses each and every element of Appellant's invention as claimed.

Appellant's arguments regarding claim 1 are directed primarily to the supposed differences between the methods disclosed by Galai and those of claim 1; in particular,

the method of URL comparison to determine whether web sites use session identifiers. Appellant argues that Galai does not disclose the limitation of claim 1, determining whether a web site corresponding to a URL uses session identifiers *...based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold* (p. 8-10 of the Brief).

Appellant argues that Galai discloses a procedure based on a comparison of two web pages, using a reduced version of the URL used to access the first web page, rather than a comparison of only the URLs without the web pages, which the arguments at p. 8 of the Brief suggest is the method claimed by Appellant in claim 1. However, a careful comparison of the limitations of claim 1, Appellant's Specification, and Galai, will reveal that both Appellant's claim 1 and Galai disclose methods to compare URLs by downloading and comparing at least two web pages corresponding to the URLs.

Appellant's Specification provides the details of how the methods claimed in claim 1 are accomplished. Notably, Appellant's Specification at p. 14, par. [0046]-[0047], which was cited in part on p. 3 of the Brief as providing support for Claim 1, discloses downloading two copies of a web page to determine whether a web site uses session identifiers, and comparing the content within the page (links) as well as the URLs to identify the pages using session identifiers. For this reason, the examiner maintains that the methods of claim 1 are unpatentable over Galai, in view of DaCosta, because Galai teaches the comparison of URLs and links from two web pages in order

to determine similarity between URLs and the corresponding web pages (Galai, p. 20, l. 10-p. 21, l. 9; p. 27; Fig. 5).

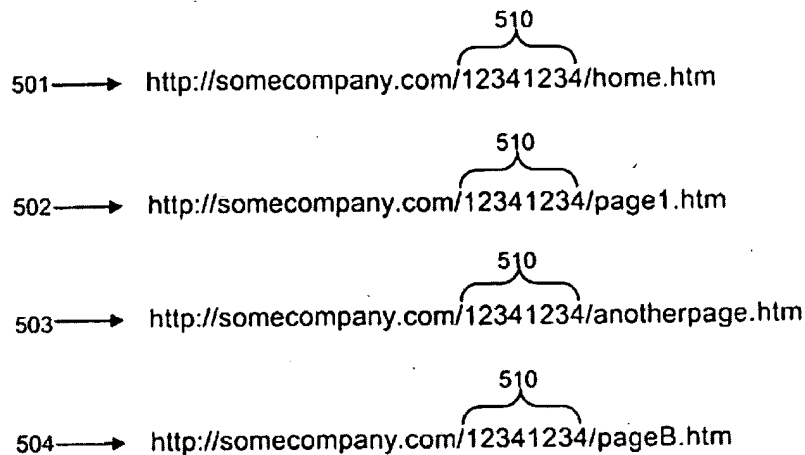


FIG. 5

Figure 5 of Appellant's Specification, at item 510, depicts a session identifier which is contained in the URLs of a site. Each URL provides an address for a different web page within the web site. The session identifier is used to track and record the activities of each individual visitor to the site, and may result in different pages and content being shown to each visitor, for example. The session identifier is a type of "redundant parameter" within the URLs, as disclosed in Galai, p. 20, l. 10-p. 21, l. 8, because it is a divisible subunit of the URL which is repeated. Galai teaches a method of comparing web pages retrieved using two different URLs, to determine the redundant parameters within the URLs, and therefore teaches a method of comparing URLs by downloading web pages (p. 27, l. 6-p. 28, l. 21).

The methods disclosed by Galai expressly teach the limitation of claim 1, *when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold*, because Galai discloses a comparison function and similarity lever which may be determined to exceed a certain threshold at p. 27, l. 14-23. While Appellant argues that the comparison function disclosed by Galai "is based on a comparison of the content of a web page" at p. 10 of the Brief, the examiner maintains that claim 1 discloses a comparison function based on a comparison of the content of a web page, as detailed in p. 14-15 of the Specification, for example. Claim 1 itself recites *receiving at least two different copies of a document associated with the URL*, and therefore is directed to comparing *URLs that change between the two different copies of the document...* and is based on a comparison of two web pages.

While Galai discloses the limitations of claim 1, in an abundance of caution, the DaCosta patent was relied upon as a secondary reference to disclose the purpose of obtaining session data for URLs. Appellant argues in regard to the DaCosta patent that DaCosta does not disclose session identifiers or determining whether a web site uses session identifiers. However, DaCosta does disclose obtaining session data for the URL at col. 4, l. 41-col. 5, l. 23. Appellant admits at p. 7, par. 4 of the Brief that "Galai discloses the existence of session identifiers" therefore the session identifiers of claim 1 are admitted prior art.

Further, as stated in the rejection of claim 1 (p. 4, par. 2 of Final Rejection), storing session data for a web site in either a URL string or a cookie was notoriously well known in the art at the time of the invention, and the examiner's assertion of official

notice, which was not traversed by Appellant, is also admitted prior art in the record (p. 4 and 15 of Final Rejection).

For these reasons, and the reasons of record, it is respectfully requested that the claim rejections of claims 1-3 and 5-8 be sustained.

"2. Claim 4." (p. 12-13 of the Brief).

While Appellant argues that Galai does not disclose the limitations of dependent claim 4, *wherein the compared URLs that change include URLs that are local to the web site*, Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20). Because Galai teaches that the method of comparing URLs may be used for any web pages, it is obvious that the URLs would include URLs local to a web site.

Appellant has not provided any special definition of the limitation "local to the web site", and because the claims must be given the broadest reasonable interpretation in light of the supporting disclosure, "local to the web site" has been interpreted as a web page of a web site.

Appellant's remaining arguments in regard to dependent claim 4 repeat the arguments set forth for independent claim 1, above, and for similar reasons as set forth for claim 1, it is respectfully requested that the claim rejections be sustained.

"3. Claims 10, 13, 14, 21, 24, and 25." (p. 14-15 of the Brief).

Appellant's arguments in regard to claims 10, 13, 14, 21, 24, and 25 repeat the arguments set forth above for independent claim 1 (and claims 2, 3, and 5-8), and for similar reasons as for claim 1, it is respectfully requested that the claim rejections of claims 10, 13, 14, 21, 24, and 25 be sustained.

"4. Claims 12, 23, and 28." (p. 15-16 of the Brief).

Appellant's arguments presented for claims 12, 23, and 28 are substantially similar to the arguments presented for dependent claim 4, above, and for similar reasons as set forth for claim 4, it is believed that the claim rejections should be sustained.

Appellant's remaining arguments in regard to claims 12, 23, and 28 repeat the arguments set forth above for independent claim 1 (and claims 2, 3, and 5-8), and for similar reasons as for claim 1, it is respectfully requested that the claim rejections of claims 12, 23, and 28 be sustained.

"5. Claims 15-18 and 20." (p. 17-18 of the Brief).

Appellant's arguments in regard to independent claim 15 state that claim 15 is directed to "a device comprising a spider component and a session identifier

component" (p. 17, par. 1 of the Brief). Appellant presents no arguments in regard to the spider component. Galai teaches a spider component at p. 4, l. 11-14.

The remaining arguments presented by Appellant in regard to claim 15 are substantially similar to the arguments presented for independent claim 1, and for similar reasons as for claim 1, it is respectfully requested that the claim rejections of claim 15 (and claims 16-18 and 20) be sustained.

"6. Claim 19." (p. 18-19 of the Brief).

Appellant's arguments presented for claim 19 are substantially similar to the arguments presented for dependent claim 4, above, and for similar reasons as set forth for claim 4, it is believed that the claim rejections should be sustained.

Appellant's remaining arguments in regard to claim 19 repeat the arguments set forth above for independent claim 1 (and claims 2, 3, and 5-8), and for similar reasons as for claim 1, it is respectfully requested that the claim rejections of claim 19 be sustained.

(11) Related Proceeding(s) Appendix

No decision rendered by a court or the Board is identified by the examiner in the Related Appeals and Interferences section of this examiner's answer.

For the above reasons, it is believed that the rejections should be sustained.

Respectfully submitted,



Amelia Rutledge

January 3, 2008

Conferees:



Lynne Browne (Appeal Practice Specialist, TQAS)



Doug Hutton (Supervisor)

DOUG HUTTON
SUPERVISORY PATENT EXAMINER